AD________________

GRANT NUMBER DAMD17-94-J-4332

TITLE: Statistical Methods for Analyzing Time-Dependent Events in Breast Cancer Chemoprevention Studies

PRINCIPAL INVESTIGATOR: George Y. C. Wong, Ph.D.

CONTRACTING ORGANIZATION: Strang Cancer Prevention Center
New York, New York 10021

REPORT DATE: October 1998

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1998	Annual (30 Sep 97 - 29 Sep 98)

4. TITLE AND SUBTITLE

Statistical Methods for Analyzing Time-Dependent Events in Breast Cancer Chemoprevention Studies

5. FUNDING NUMBERS

DAMD17-94-J-4332

6. AUTHOR(S)

George Y.C. Wong, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Strang Cancer Prevention Center
New York, New York 10021

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

19990318 073

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The overall aim of our research proposal is the statistical nonparametric inference of the redistribution-to-the-center estimator (RTCE) and the generalized maximum likelihood estimator (GMLE) for the survival function of a time-to-event variable that is subject to interval censoring. The RTCE, which is proposed by us, has a closed-form expression and is equal to the GMLE under a homogeneous condition. The GMLE is the standard optimal estimator in survival analysis. However, it cannot be expressed in a closed-form expression, and asymptotic distribution theory for it has been limited. From the asymptotic study of the RTCE, we have gained important insight into proofs of asymptotic properties of the GMLE. In the past four years we have established consistency, asymptotic normality and asymptotic efficiency of the GMLE under a variety of conditions. In addition, we have derived an asymptotic nonparametric two-sample distance test procedure for comparing two populations. Finally, we have begun investigating the asymptotic inference of Cox regression model for interval-censored data by establishing consistency and asymptotic normality of the GMLE of the regression parameters under some finite assumptions. These preliminary results are being applied to a breast cancer prognostic relapse follow-up study.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer, Interval Censorship, Consistency, Asymptotic Normality, Asymptotic Efficiency	15
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18 298-102

USAPPC V1.00

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

George Wong 10/28/98

PI - Signature Date

A. TABLE OF CONTENTS

B. INTRODUCTION

In clinical follow-up studies, subjects are monitored at regular time intervals for a physical condition. It is often the case that an event under observation can take place in between two successive visits, and it may not be possible for the subject to know the time to such an event exactly. For example, consider the situation in which a group of women at high risk for breast cancer is asked to take a chemopreventive substance for a fixed time period. At the end of the period, each participating woman is required to submit a blood or urine sample at regular intervals in order to monitor the level of a validated intermediate biomarker. Let X denote the time from cessation of use of the agent to the loss of its protective effect, quantified as a return to baseline value of the biomarker. If a woman submits a sample for assay on a daily basis, the value of X can be observed exactly, unless the protective effect is still present by the time the study is terminated so that X is right censored in the usual sense of survival analysis. In practice, however, the follow-up interval can be a week or longer; therefore the exact value of X is generally unknown but is known to lie between the time points L and R, where L is the number of days from cessation of agent intake to the last time the sample was assayed and the protective effect was still present, and R is the number of days from cessation of agent intake to the most recent time the sample was assayed. If the protective effect is still present, then R takes the value infinity. In any case, when the value of X is only known to lie between (L, R), we say that X is censored in the interval (L, R). Therefore the observed data consist of either censoring intervals (L, R) or exact observations $X = L = R$.

Our research project is concerned with nonparametric estimation of the distribution function $F(t) = Pr(X \leq t)$ of a real-valued random variable X, or equivalently its survival function $S(t) = 1 - F(t)$, when the sample data are incomplete due to restricted observation brought about by interval censoring. Generalized maximum likelihood (GML) method in the sense of Kiefer and Wolfowitz [1] is the standard practice of estimating S. At present, there are two iterative computation procedures that will yield the GML estimate (GMLE) of S at convergence. The first one is due to Peto [2] and makes use of the Newton's method. The second is due to Turnbull [3] and makes use of a simpler but slower algorithm called self-consistent algorithm. A solution to this algorithm is also called a self-consistent estimator (SCE).

Because there is no closed-form expression for the GMLE of S, it has been difficult to study its asymptotic statistical properties, including consistency, normality and efficiency. Such a setback in the statistical development of the GMLE has severely limited its use in the statistical analysis of interval-censored (IC) data.

Before we began our funded Army research, we had extended Efron's redistribution-to-the-right idea for right-censored data [4] and proposed a redistribution-to-the-center (RTC) method to yield a nonparametric estimator of S which are called RTC estimate (RTCE).

Such an estimator has a closed-form expression and can be readily calculated for IC data of any dimension. IC data are said to satisfy DI (disjoint or included) condition if for every two censoring intervals, either they are disjoint or one is a subset of the other. For instance, in a clinical study in which every subject has the same follow-up schedule, say at time point a_1, a_2, ..., a_k, then $\{L, R\} = \{0, a_1\}$, or $\{a_i, a_{i+1}\}$ or $\{a_i, \infty\}$. A sample of such IC data $\{L_1, R_1\}, ..., \{L_n, R_n\}$ will satisfy DI condition. We had shown that under DI condition, RTCE is actutally GMLE itself. This important observation, together with the availability of an explicit expression, had motivated us to submit the present proposal on RTCE to the Army.

In our first year of research, we completed our research for Task 1 and Task 2 in the Statement of Work for RTCE. However, we also discovered that in the case of non-DI data, RTCE may be different from GMLE, and RTCE is not always consistent. The interesting and intriguing observation is that the difference between RTCE and GMLE is small, at least based on our limited simulation studies [5]. In establishing consistency result for RTCE under DI condition, we had gained important insight into proofs of asymptotic properties for GMLE, which does not possess a closed-form expression. Because GMLE is the preferred estimator for S, we decided to focus our attention on GMLE instead of RTCE for the remainder of the funded research, and we have successfully completed all the tasks stated in the Statement of Work for GMLE.

C. BODY

C.1. Basic setup

Interval-censored data can arise in the following four situations:

1. Case 2 IC data (C2 data) consist of right-censored ($R = \infty$), left-censored ($L = 0$) and strictly interval-censored observations ($0 < L < R < \infty$). These are by far the most common type of IC data in clinical follow-up studies.
2. Mixed IC data (MIC data) consist of both C2 data and exact observations ($L = R$). Yu, Li and Wong [6] presented an example involving MIC data from a breast cancer follow-up study.
3. Case 1 IC data (C1 data)) consist of either right-censored or left-censored observations. For example, when an animal is sacrificed for inspection of a tumor formation, time to appearance of the tumor is C1 interval censored. Examples of C1 data can be found in [7] and [8].
4. Doubly-censored data (DC data) consist of right-, left-censored and exact observations. An example with DC data is given in [9].

We have formulated four different interval censorship models corresponding to the four IC data types. To study the asymptotic properties of the GMLE, we make use of the following assumptions:

(AS1) The censoring distribution is discrete but the survival distribution is arbitrary.

(AS2) The support set of the censoring vector is finite, but the survival distribution is arbitrary.

(AS3) A probability restriction. See Section C.

(AS4) A probability restriction. See Section C.

(AS5) The censoring distribution and the survival distribution are arbitrary, but have to satisfy some regularity conditions, stated in Gu and Zhang [10].

C.2. Case 1 model

Case 1 model for C1 data assumes that the survival time X and a random inspection time Y are independent. We always observe Y. However, X is not fully observed except that we know that either $X \leq Y$ or $X > Y$. Under assumption AS1, we have shown that GMLE is strongly consistent, asymptotically normal and asymptotically efficient at all the inspection times. The results are published in Yu, Schick, Li and Wong [11].

C.3. Case 2 model

The C2 model for C2 data assumes that X and the random censoring vector (Y, Z) are independent and that $Y < Z$ with probability one. We do not observe X except that we know X is before Y, or between Y and Z, or after Z. We state an assumption for C2 model as follows:

(AS3) $\mathrm{P}\{X \in I_i \cap I_j\} > 0$ for any two realizations of (L, R), $(L_i, R_i) = I_i$ and $(L_j, R_j) = I_j$, provided $I_i \cap I_j \neq \emptyset$.

Under the assumption AS1, we have shown that GMLE is strongly consistent. Under the assumptions AS2 and AS3, we have shown that GMLE is asymptotically normal and efficient. The results are published in Yu, Schick, Li and Wong [12].

C.4. MIC model

Mixture interval censorship (MIC) model for MIC data assumes that an IC observation is drawn from a probability mixture of C2 model and the usual right censorship model for right-censored data.

Define $\tau = \sup\{t;\ Pr(\min(X,T) \leq t) < 1\}$, $\tau_Y = \sup\{t;\ Pr(Y \leq t) = 0\}$. and $\tau_Z = \sup\{t;\ Pr(Z \leq t) < 1\}$. We assume that $\tau \geq \tau_Z$. We state an assumption for MIC model as follows:

(AS4) $\mathrm{Pr}(L = \tau) > 0$ if $Pr(X < \tau) < 1$ and $\mathrm{Pr}(R = \tau_Y) > 0$ if $Pr(X \leq \tau_Y) > 0$.

Under assumptions AS2 and AS4, we have shown that GMLE is strongly consistent (Yu, Li and Wong [6]), and under assumptions AS2, AS3 and AS4, GMLE is asymptotically normal (Yu, Li and Wong [13]). Recently, we have been able to establish these asymptotic properties without the need of assumption AS2. A manuscript on these results has been submitted for publication (Yu, Li and Wong [14]).

C.5. DC model

The DC model for DC data assumes that X and a random vector (Y, Z) are independent and $Y < Z$ with probability one, and that X is uncensored if $Y < X \leq Z$, right censored if $Z < X$ and left censored if $X \leq Y$. Let S_Z and S_Y be the survival functions of Z and Y, respectively, and let $K = S_Y - S_Z$. We state an assumption for DC model as follows:

(AS5) $\qquad K(x-) > 0$ for all x such that $S(x) < 1$ and $S(x-) > 0$,

We have shown in a submitted manuscript (Yu and Wong [15]) that in order to establish asymptotic results, GMLE has to be modified. Under assumptions AS4 and AS5 we have shown that the modified GMLE is strongly consistent and is asymptotically normal and efficient under assumptions AS3, AS4 and AS5.

C.6. Two-sample nonparametric test

Based on the asymptotic results that we have established for different IC models, we have successively derived the asymptotic distribution of the following two-sample distance test statistics for each model:

$$D = \int_{\tau_1}^{\tau_2} W(t)(\hat{S}_1(t) - \hat{S}_2(t))dt,$$

where τ_1 and τ_2 are specified time point and $W(t)$ is a weight function. A manuscript on the asymptotic results of D is being submitted for publication (Wong and Yu [16]).

C.7. Proportional hazards model

In our original proposal, we had assigned three months of time for Task 7 on Cox regression for IC data. However, we have realized that statistical inference for the parameter $\underline{\beta}$ in Cox regression under interval censorship is much more involved than its counterpart in the usual right-censored situation. In the latter case, the maximum likelihood estimator (MLE) of $\underline{\beta}$ does not depend on the baseline survival function $S_0(t)$ owing to the simple nature of the partial likelihood. However, such simplicity of likelihood function does not carry over to the interval censorship model, and maximum likelihood estimation of $\underline{\beta}$ will involve GML estimation of $S_0(t)$ at the same time, thus resulting in a difficult high-dimensional estimation problem.

Under the restrictive assumption that both X and the censoring vector take on finitely many values, we have proved that the MLE of $\underline{\beta}$ and the GMLE of $S_0(t)$, and hence the survival function $S(t|\underline{Z}) = S_0(t)exp^{\underline{\beta}'\underline{Z}}$, where $\underline{Z}$ denotes a vector of covariates for Cox regression, are consistent and asymptotically normal (Li, Yu and Wong [17]). Much more effort is needed to pursue research on the asymptotic inference of Cox regression model under more relaxed assumptions on the distributions of X and the censoring vector.

C.8. Computer software

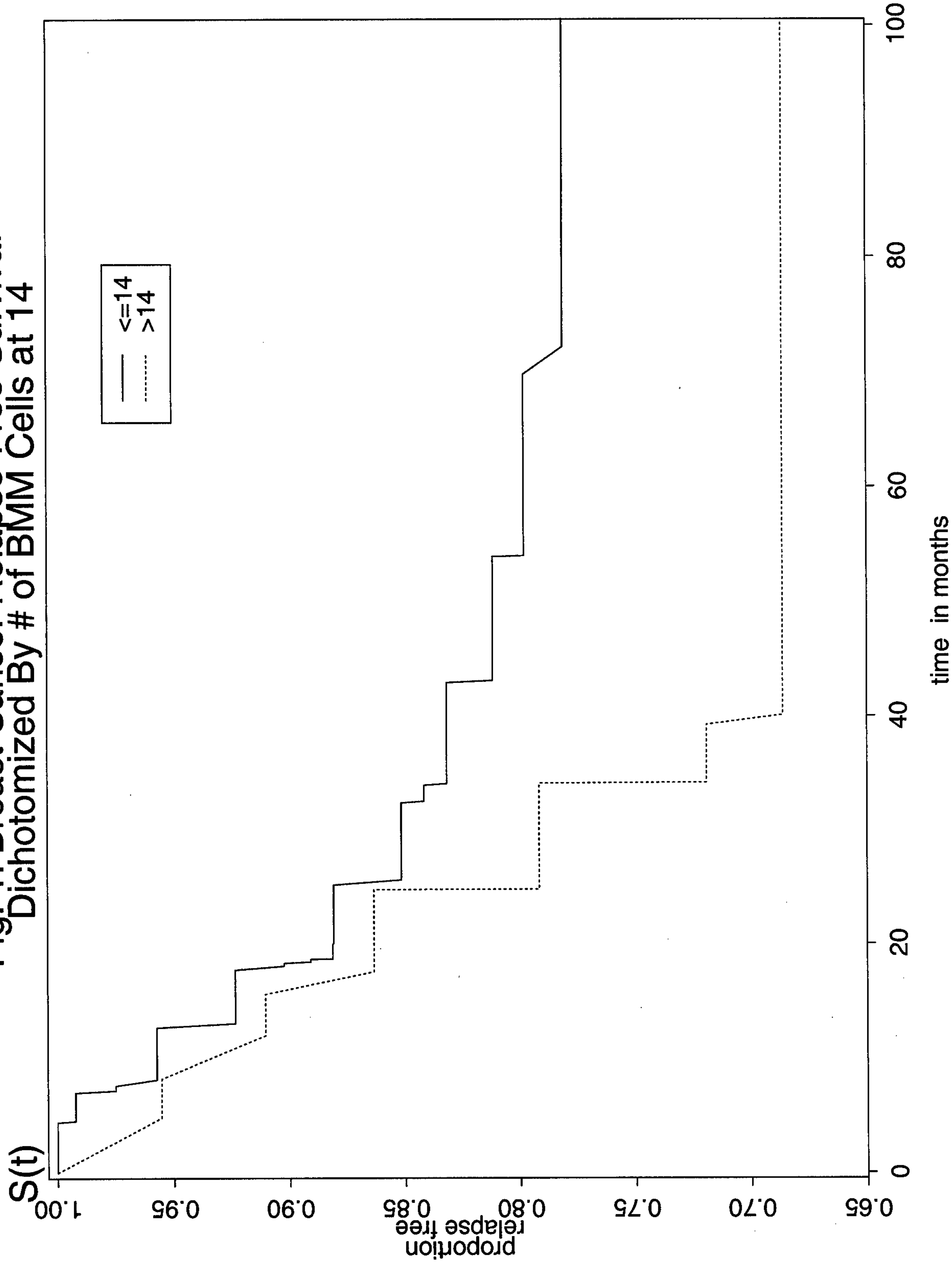
We have made it available to the public a set of computer programs for calculating RTCE and GMLE, for carrying out asymptotic inference of GMLE for all patterns of interval censorship, and for evaluating the Z-score of the proposed two-sample weighted distance test. These programs can be accessed via the internet at qyu@math.binghamton.edu.

C.9. Applications to breast cancer research

We have applied our results on asymptotic inference of GMLE for C2 model to two breast cancer research projects. The first project is concerned with a chemoprevention intervention trial of indole-3-carbinol (I3C) for breast cancer which is being conducted at Strang Cancer Prevention Center. The statistical question of interest is the estimation of duration of sustaining effect of I3C, which is C2 censored. A preliminary report on a short-term trial has recently been published [18]; however, a longer trial lasting for more than one year is still ongoing so that more informative data on duration of sustaining effect can be obtained.

The second project is a standard breast cancer relapse follow-up study based on data from 374 women with stages I - III unilateral invasive breast cancer surgically treated at Memorial Sloan-Kettering Cancer Center between 1985 and 1990. The median follow-up duration was 46 months. Relapse time was given by the time interval between surgery and the initial relapse. A relapse that took place between two successive follow-up visits was regarded as interval censored. If a patient did not relapse towards the end of the study, then her relapse time was right censored. Of the 374 observations, 300 were right censored (no relapse), 21 were left censored and 53 were strictly interval censored (74 relapses). Bone marrow micrometastasis (BMM) was determined for each woman at the time of surgery. An important question is whether remission duration is related to the extent of initial tumor burden defined as number of BMM cells detected. Figure 1 compares the relapse-free GMLE curves of patients with number of BMM ≤ 14 versus those with number of BMM > 14. Our asymptotic two-sample distance test yielded a P value close to 0.1. An abstract on a detailed prognostic analysis of the entire data set using our asymptotic results on C2 data has been accepted for presentation at the annual San Antonio Breast Cancer Symposium in December 1998.

Fig. 1. Breast Cancer Relapse-Free Survival
Dichotomized By # of BMM Cells at 14



D. CONCLUSIONS

In the four years of our Army grant, we have successfully completed our research objectives on the asymptotic inference of the GMLE of the survival function under different interval censorship models, including consistency, asymptotic normality and asymptotic efficiency. The results which we have established provided clinicians and basic science researchers in breast cancer with a set of fundamentally important statistical tools for the analysis of all types of interval-censored data (C2, MIC, DC and C1 data) that are encounter in breast cancer research. We have also made available to the general public a set of computer programs for carrying out the asymptotic generalized maximum likelihood inference procedure for all types of interval-censored data.

E. REFERENCES

[1] Kiefer, J and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* 27, 887-906.

[2] Peto, R. (1973). Experimental survival curves for interval-censored data. *Appl. Statist.* 22, 86-91.

[3] Turnbull, B. W. (1976). The empirical distribution function with arbitrary grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser.* B, 38, 290-295.

[4] Efron, B (1967). The two sample problem with censored data. *Fifth Berkeley Symposium on Mathematical Statistics.* University of California Press, 831-853.

[5] Wong, G. Y. C. and Yu, Q. (1995). Estimation of a survival function with interval-censored data; A simulation study. (Preprint).

[6] Yu, Q., Li, L. and Wong, G. Y. C. (1997). On consistency of the self-consistent estimator of survival functions with interval censored data. *Scan. J. of Statist.* (Accepted).

[7] Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T. and Silverman, E. (1955). An empirical distribution function for sampling incomplete information. *Ann. Math. Statist.* 26, 641-647.

[8] Keiding. N. (1991) Age-specific incidence and prevalence: A statistical perspective (with discussion) *JRSS*, A. 154, 371-412.

[9] Leiderman, P.H., Babu, D., Kagia, J., Kraemer, H.C. and Leiderman, G.F. (1973). African infant precocity and some social influences during the first year. *Nature*, 242, 247-249.

[10] Chang, M.N. and Yang, G. (1987). Strong consistency of a self-consistent estimator of the survival function with doubly censored data. **Ann. Statist.** 15, 1536-1547.

[11] Yu, Q., Schick, A., Li, L. and Wong, G. (1998). Asymptotic properties of the GMLE in the case 1 interval-censorship model with discrete inspection times. *Canadian Journal of Statistics.* Vol. 4.

[12] Yu, Q., Schick, A., Li, L. and Wong, G. Y. C. (1998). Asymptotic properties of the GMLE of a survival function with case 2 interval-censored data. *Prob. & Statist. Letters.* 37, 223-228.

[13] Yu, Q., Li, L. and Wong, G. Y. C. (1998). Asymptotic variance of the GMLE of a survival function with interval-censored data. *Sankhya,* A. (Accepted).

[14] Yu, Q., Li, L. and Wong, G. Y. C. (1998). Asymptotic properties of NPMLE with mixed interval-censored data. (Submitted to Annals of Statistics.)

[15] Yu, Q., and Wong, G. (1998). A modified GMLE with doubly-censored data. (Submitted to *Australian J. of Statist.*).

[16] Wong, G. Y. C. and Q. Yu (1998). A generalized two-sample distance test with interval-censored data. (Submitted to *Statistics in Medicine*).

[17] Li, L., Yu, Q. and Wong, G. Y. C. (1998). Proportional hazards model with interval-censored and exact observations. (Under preparation).

[18] Wong, G. Y. C., Bradlow, H. L., Sepkovic, D., Mehl, S., Mailman, J. and Osborne, M. P. (1997). A dose-ranging study of indole-3-carbinol for breast cancer prevention. *Journal of Cellular Biochemistry Supplements* 28/29, 111-116.

F. APPENDICES

1. List of publications (9):

[a] Li, L., Watkins, T. and Yu, Q. (1997). An EM algorithm for smoothing the self-consistent estimator of survival functions with interval-censored data. *Scan. J. Statist.* 24, 531-542.

[b] Yu, Q., Li, L. and Wong, G. Y. C. (1998). On consistency of the self-consistent estimator of survival functions with interval censored data. *Scan. J. of Statist.* (Accepted).

[c] Yu, Q., Schick, A., Li, L. and Wong, G. Y. C. (1998). Asymptotic properties of the GMLE in the case 1 interval-censorship model with discrete inspection times. *Canadian Journal of Statistics.* Vol. 4.

[d] Yu, Q., Li, L. and Wong, G. Y. C. (1998). Asymptotic variance of the GMLE of a survival function with interval-censored data. *Sankhya,* A. 60, 184-197.

[e] Yu, Q., Schick, A., Li, L. and Wong, G. Y. C. (1998). Asymptotic properties of the GMLE of a survival function with case 2 interval-censored data. *Prob. & Statist. Let.* 37, 223-228.

[f] Yu, Q. and Wong, G. Y. C. (1998). Consistency of self-consistent estimators of a discrete distribution function with bivariate right-censored data. *Commun. Statist.* 27, 1461-1476.

[g] Wong, G. Y. C. and Yu, Q. (1998). Generalized MLE Of a joint distribution function with multivariate interval-censored data. *J. of Multi. Anal.* (Tentative accepted).

[h] Schick, A. and Yu, Q. (1998). Consistency of the GMLE with mixed case interval-censored data. *Scan. J. Statist.* (Accepted).

[i] Li, L. and Yu, Q. (1997). Self-consistent estimators of survival functions with doubly-censored data. *Commun. Statist.*, 2609-2623.

2. List of manuscripts submitted (3):

[j] Yu, Q., Li, L. and Wong, G. Y. C. Asymptotic properties of NPMLE with mixed interval-censored data. (Submitted to Annals of Statistics.)

[k] Yu, Q. and Wong, G. Y. C. A modified GMLE with doubly-censored data. (Submitted to *Australian J. Statist.*).

[l] Wong, G. Y. C. and Yu, Q. A generalized two-sample distance test with interval-censored data. (Submitted to *Statistics in Medicine*).

3. List of manuscripts under preparation (3):

[m] Li, L., Yu, Q. and Wong, G. Y. C. Proportional hazard model with interval-censored and exact observations.

[n] Wong, G. Y. C. and Yu, Q. Strong consistency of the generalized MLE of a survival function under the DI model.

[o] Wong, G. Y. C. and Yu, Q. Estimation of a survival function with interval-censored data under the DI model.

4. List of conferences attended

(1) *Institute of Mathematical Statistics Meeting no. 246 8-12 July 1996 Sydney, N.S.W.*

TITLE: Variance of the MLE of a Survival Function with Interval Censored Data

Qiqing /Yu , Linxiong /Li and George Y. C. /Wong

SUNY at Binghamton, University of New Orleans and Strang Cancer Preventive Institute

ABSTRACT: Interval-censored data consist of n pairs of observations (l_i, r_i), $i = 1, ..., n$, where $l_i \leq r_i$. We either observe the exact survival time X if $l_i = r_i$ or only know $X \in (l_i, r_i)$ otherwise. We established the asymptotic normality of the nonparametric MLE of a survival function $S(t)$ $(= Pr(X > t)$ with such interval-censored data and present an estimate of the asymptotic variance of the MLE. We show that the convergence rate in distribution is in $\sqrt{n}$. Simulation study also supports our result. An application to the cancer research is presented.

Paper presented in person, contributed paper.

(2) *Institute of Mathematical Statistics Meeting no. 247 4-8 August 1996 Chicago, Illinois*

TITLE: ESTIMATION OF A SURVIVAL FUNCTION WITH CASE 1 INTERVAL -CENSORED DATA

Qiqing /Yu , Anton /Schick, Linxiong /Li and George Y. C. /Wong

SUNY at Binghamton, University of New Orleans and Strang Cancer Preventive Institute

ABSTRACT: Case 1 interval censored data consists of either right-censored data or left-censored data but not exact observations. Let $F(x)$ and $G(y)$ be the distribution functions of the survival time X and censoring time Y, respectively. Groeneboom and Wellner (1992, p.100) (G&W) establish the consistency and asymptotic distribution of the MLE of F under the assumption that $F'(x)$ and $G'(y)$ are both positive and continuous. Under the assumption that X *is arbitrary, but* Y *takes on finitely many values*, we establish the consistency, asymptotic normality and efficiency of the MLE of $F(y)$ at the observations, y, of Y, and present a consistent estimate of the asymptotic variance of the MLE. The convergence rate in distribution is $n^{1/3}$ under G&W's assumption, but it is $\sqrt{n}$ under our assumption. Simulation results indicates that the sample variance is very close to the theoretical value of the asymptotic variance given in our paper, even for a sample size of 100.

Paper presented in person, contributed paper.

(3) ICSA 1997 Applied Statistics Symposium
May 30 - June 1, 1997
Rutgers University, New Jersey, USA
Title: Asymptotic Properties Of Self-Consistent Estimators of A Survival Function
by Qiqing Yu and George Y. C. Wong.
SUNY at Binghamton and Strang Cancer Preventive Institute

ABSTRACT: The asymptotic properties of the nonparametric maximum likelihood estimator and other estimators of a joint distribution function F of a bivariate random vector X with right-censored data have been studied by several authors. Among others, an important assumption made in their studies is that X lives on a rectangle region $[0, a] \times [0, b]$ which can be observed. However, in many follow-up studies, $a = b = L$ is the length of the study period and X lives on a region larger than $[0, L] \times [0, L]$. Thus it is of interest to study whether the asymptotic results established by these authors are still valid without that restriction. In this direction, we established the strong consistency of self-consistent estimators of a discrete distribution function.